NOT TO BE FURTHER COPIED OR REPRODUCED

(* BATCHES *)

2

Easy Summaries

(Ranks and Depths)

		Exh.	Page
Introduction 2-1		a	
		b	
2A.	Extremes and median 2-2	1	2-5
		2	2-7
2B.	Counting-in 2-3	c	
	rank, defined at 2-3	3	2-10/11
	depth, defined at 2-4	d	
	interpolated ranks, defined at 2-6	4	2-14/15
		5	2-19
2C.	Wings and hinges 2-6	e	
	wings, defined at 2-8	f	
	hinges, defined at 2-8	g	
		h	
2D.	5-number summaries 2-13	k	
		l	
2E.	Some other convenient numbers 2-20	p	
	midspread, defined at 2-20	6	2-34
	range, defined at 2-20	7	2-35
	trimean, defined at 2-21	8	2-36
2F.	Pentameans? 2-23		
2G.	Halving and trimeans (helpful) 2-25		
	**********		
2H.	How far have we come? 2-27		
	**********		
2K.	Middle means 2-30		
	middle mean, defined at 2-30		
	midmean, defined at 2-31		
2L.	Why might the trimean behave like a middle mean? (optional background) 2-32		
	**********		
2P.	Problems for chapter 2		

2

Easy Summaries
(Ranks and Depths)

We have learned to scratch down batches. Stem-and-leaf displays let us do this easily and rapidly. Even more importantly, they let us look at the general pattern of the batch.

When this pattern is distinctive, we can easily say so in words. Exhibit 10 of Chapter 1, for example, shows a pattern of three quite distinct groupings. The patterns of Exhibits 4 and 5 of that chapter, for another, stop much more abruptly toward smaller values than toward larger ones.

We would like to be able to summarize the most frequently occurring characteristics of the pattern of a batch. It will be good to do this in terms of a few numbers that are easily understood and to agree on what these numbers are, what they are called, how they are marked, and how they are routinely written down. To do this will be the central aim of this chapter.

To do this well, we must limit our objectives. It would be wrong to expect a standard summary to reveal the unusual -- no matter how important the unusual may be when it occurs. Things like the separation of the pattern into groupings -- illustrated by Exhibit 10 of Chapter 1 -- ought not to show in an easy summary of routine form. On the one hand, it deserves words. On the other, trying to get it in would complicate our procedures unduly and confuse the results for all the many batches where nothing of the sort happens.

We ought to depend on stem-and-leaf displays -- or other visual devices -- to show us the unexpected. Standard numbers should be to reveal the more or less expected.

While we are at it, we shall go a step further, and develop some summaries of even more focussed aim.

2A. Extremes and Median

If we are to select a few easily-found numbers to tell something about a batch as a whole, we certainly want:

- the extremes -- the highest and lowest values -- which we will mark "*".
- a middle value.

We would like these values to be easy to write down, whether the total count of the batch is 8, or 82, or 208.

Once we have the batch in stem-and-leaf form we have its values almost in order. Once we have the check count we can count up or down a stem at a time. The easy values to find are those that are fixed by counting.

If the batch has an even number of values, there are two in the middle by count. If there are 50 in all, halfway from 1 to 50 is 25.5, so that the 25th and 26th are the middle pair.

If the count is an odd number, there is a single value in the middle by count. If there are 13 in all, halfway from 1 to 13 is 7, so the 7th is it.

The median is defined to be

$$\text{median} = \begin{cases} \text{the single middle value} \\ \text{the mean of the two middle values} \end{cases}$$

We shall mark it "M".

For Exhibit 10 of Chapter 1, we find these values

$$\begin{array}{r|r} * & 3 \\ M & 46 \\ * & 203 \end{array} \quad \text{(hundreds of feet)}$$

In this case, the extremes are easiest. The total count was 50, making the rank of the median 25.5. The two numbers to which ranks 25 and 26 are assigned were 44 and 48, so the median was

$$46 = \frac{1}{2}\,(44+48)$$

2B. <u>Counting-in</u>

To get an easy number in the middle of the batch -- assuming that the batch is in stem-and-leaf form -- we count in from either end, starting by counting the extreme value "1". This procedure assigns a

rank

to each number in the batch.

We can start from either extreme. In such phrases as "he was first in his class" it seems natural to start by ranking the highest value as "1", to rank down . Many times, by contrast, we like to have the rank increase as the value increases, to rank up . This calls for ranking the lowest value "1".

The least rank -- either up or down -- that we can give to an observation is its

depth

The depth of an extreme is always 1.

Whichever way we rank a batch of 17, the median is rank 9. 9 is what $\frac{1}{2}$ (17) "grows to" in the sense of the introduction to chapter 1. So long as the total count of a batch is an odd number, the rank of the median is what $\frac{1}{2}$ (total count) grows to.

In a batch of six, say for example

7
12
18
20
27
54

exhibit 1 of chapter 2

The two directions of ranking illustrated
(Data of exhibit 3 of chapter 1)

Upward Rank	Prices	Downward Rank	Depth
1	150	17	1
2	250	16	2
3	688	15	3
4	695	14	4
5	795	13	5
6	795	12	6
7	895	11	7
8	895	10	8
9	895	9	9
10	1099	8	8
11	1166	7	7
12	1333	6	6
13	1499	5	5
14	1693	4	4
15	1699	3	3
16	1775	2	2
17	1895	1	1

Remark:

upward rank + downward rank = 1 + total count

for each and every entry. This is a general property.

The median is the mean of the two center values

$$\frac{1}{2}(18+20) = 19$$

The ranks of these values are 3 and 4, no matter which end we rank from. Half-way between rank 3 and rank 4 should come rank 3.5.

Now 3.5 is what $\frac{1}{2}$ (6) grows to, according to the original definition. The same sort of thing happens for batches of any even count. If we do the natural thing, and use half-integer

interpolated ranks

for the midvalues half-way between the values to which we have given adjacent integer ranks, we will have the same expression

$$\text{rank of median} = \text{what } \frac{\text{count}}{2} \text{ grows to}$$

for batches of all counts, odd or even.

Exhibit 2 shows some examples.

Using interpolated ranks and values and the concept of "growing to" has made it easy to describe the median. The same ideas can make it easy to describe other useful quantities.

2C. Wings and Hinges

We have grasped 3 numbers to define the ends and the middle of a batch. These are useful, but are rarely enough. The easy way to pick out a few more numbers for a summary is

exhibit 2 of chapter 2

Examples of interpolated ranks

A) In the DATA of EXHIBIT 1

Interpolated Rank	Value and Its Origin
1.5	200 = $\frac{1}{2}$ (150+250)
2.5	469 = $\frac{1}{2}$ (250+688)
3.5	691.5 = $\frac{1}{2}$ (688+695)
4.5	745 = $\frac{1}{2}$ (695+795)
5.5	795 = $\frac{1}{2}$ (795+795)

(and so on)

B) In the BATCH of 6

Rank		Interpolated	
Rank	**Number**	**Number**	**Rank**
1	7		
		9.5	1.5
2	12		
		15	2.5
3	18		
		19	3.5
4	20		
		23.5	4.5
5	27		
		40.5	5.5
6	54		

to use depths. A natural thing to do, since we have already used

$$\frac{\text{count}}{2}$$

is to use

$$\frac{\text{count}}{4}$$

which will be crudely half-way from the extremes to the median.

Together with the median, the two quantities thus defined will break the batch up into four parts of rather similar count -- four quarters. We need a name for the end quarters thus formed. Germans sometimes call them "die fliegende Viertel" -- the flying quarters. For shortness we will call them

wings.

We will refer to the points that mark off the wings, whose depth is

what $\frac{\text{count}}{4}$ grows to

very frequently. If the wings are thought of as flapping, the points ought to be

hinges.

Hinges we shall call them, and "H" we shall mark them.

Exhibit 3 shows one form of display that emphasizes the way median and hinges break up the batch into quarters. In panel B, this form of display is combined with a mixed-leaf stem-and-leaf display. The combination takes a moment's thought, but leads us to a clearer picture of this batch of 35 ultimate powers of hydroelectric plants.

Panel C shows what happens when the total count is a multiple of 4: the hinges are interpolated, not observed. (Of course this is just what happens with the median even more often -- every time the total count is a multiple of 2.)

exhibit 3 of chapter 2

Wings and hinges illustrated

A) The 17 PRICES of EXHIBIT 1

	Median	Middle	Hinges	Wings	Extremes
					150
				250	
				688	
				695	
			795		
		795			
		895			
		895			
	895				
		1099			
		1166			
		1333			
			1499		
				1693	
				1699	
				1775	
					1895
(Counts)	(1)	(3+3)	(1+1)	(3+3)	(1+1)

exhibit 3 (Cont'd)

B) The 35 ULTIMATE POWERS of EXHIBIT 6 of CHAPTER 1

	Median	Middle	Hinges	Wings	Extremes
1				6	5
2				2445	
3		66	3	02	
4		058			
5		0			
6		0			
7	5	1			
8					
9					
10		08			
11		4			
12		00			
$\overline{1}$		34,50,	62,		
2				25,50,85,	
3				79,	
4				24,	
$\overline{0}$				900,	
1				345,	974,
(Counts)	(1)	(8+7)	(1+1)	(7+7)	(1+1)

Extremes: 15 and 1974

Hinges: 33 and 162

Median: 71

exhibit 3 (Cont'd)

C) What HAPPENS to BATCHES of 8 -- or of 4, 12, 16, 20,...

	Middle	Wings
Median	**(Hinges)**	**Extremes**
		214
		207
	(205)	
	203	
	198	
(195.5)		
	193	
	191	
	(188)	
		185
		184
(Count) (none)	(2+2) (none)	(1+1) (1+1)

Notes: 1) Since 8/4 grows to 2.5, the hinges are interpolated, not observed. This happens for every multiple of 4.

2) Since 8/2 grows to 4.5, the median is interpolated, not observed. This happens for every even count.

2D. 5-Number Summaries

Extremes, hinges, and median together make up a convenient 5-number summary.

Exhibit 4 shows some examples of such summaries. In the first two, we can count in to our desired ranks without need for any special aid, as we usually can for smallish batches.

In panels D and E, while we could count in just by pointing at entries in the stem-and-leaf in succession, this is likely to be an effort. To make things easy -- and thereby less subject to error -- we have added a new column to show which group of depths goes with each stem. It is now easy to pick out the stem that contains our answer. Once picked out, we need only finish our count within it. Panel E shows this column where and how we usually want it.

Though it is somewhat concealed, the tendency of the data in panel D to trail out further toward high values can be seen in the summary

[82,*HMH*: 38,45,57,69,94]

The hinges, 45 and 69, are each 12 away from the median, which is quite symmetrical, to be sure. The extremes, however, are, respectively, 7 and 25 units outside the hinges, which shows a considerably longer trail off toward higher values.

exhibit 4 of chapter 2

Some 5-number summaries

A) The BATCH of 17 in EXHIBIT 3

$$\frac{\text{count}}{2} = \frac{17}{2} \text{ grows to } 9$$

$$\frac{\text{count}}{4} = \frac{17}{4} \text{ grows to } 5$$

median = 895

hinges = 795 and 1499

extremes = 150 and 1895

B) CONDENSED NOTATIONS -- two alternative forms, vertical and horizontal

```
17|
 *|150
 H|795
 M|895
 H|1499
 *|1895
```

[17,*HMH*: 150,795,895,1499,1895 dollars]

Notes: 1) We use the form that fits best into the place where it is to be used.

2) Both forms include the total count, here 17, of the batch. It pays to be reminded of this.

exhibit 4 (Cont'd)

C) The BATCH of 8 in PANEL C of EXHIBIT 3

$$\frac{\text{count}}{2} = \frac{8}{2} \text{ grows to } 4.5$$

$$\frac{\text{count}}{4} = \frac{8}{4} \text{ grows to } 2.5$$

21	4
20	37
19	138
18	45

median = 195.5

hinges = 188 and 205

extremes = 184 and 214

[8,*HMH*: 184,188,196, 205,214]

Note rounding of 195.5 to 196. This is a matter of judgement.

D) The BATCH of 82 in EXHIBIT 4 of CHAPTER 1

$$\frac{\text{count}}{2} = \frac{82}{2} \text{ grows to } 41.5$$

$$\frac{\text{count}}{4} = \frac{82}{4} \text{ grows to } 21$$

exhibit 4 (Cont'd)

		Values	Depth
(1)	3	8	1 to 1
(19)	4	0001111122222223344	2 to 20
(12)	4	555566678899	21 to 32
(6)	5	001124	33 to 38
(12)	5	577778899999	✓ 12 — 39 to 50 / 33 to 44
(6)	6	112444	27 to 32
(6)	6	588899	21 to 26
(9)	7	001223334	12 to 20
(5)	7	56678	7 to 11
(3)	8	033	4 to 6
(1)	8	8	3 to 3
(2)	9	24	1 to 2

Depth 41.5 is among depths 39 to 50, hence among the 50's. Counting upward "5 is 39, 7 is 40, 7 is 41, 7 is 42,..." fixes the median at 57.

Depth 21 up is among depths 21 to 32, hence among the 40's. Counting upward "5 is 21,..." fixes one hinge at 45.

Depth 21 down is among depths 21 to 26, hence among the 60's. Counting downward "9 is 21,..." fixes the other hinge at 69. Thus we find

[82,*HMH*: 38,45,57,69,94 tens of square miles]

Notice the check now takes place where the two rankings need a single value. Here we have 38+12+32 = 82, which checks.

exhibit 4 (Cont'd)

E) The 50 VALUES of EXHIBIT 10 of CHAPTER 1 -- with depths

5	0	34588
11	1	236788
17	2	003448
to 22	3	24556
✓ 5	4	01489
to 23	5	03347
18	6	367
15	7	2
14	8	8
.	9	
.	10	
13	11	2
12	12	678
9	13	12588
4	14	445
..	15	
..	16	
.	17	
.	18	
.	19	
1	20	3

[50,*HMH*: 3, 20, 46, 112, 203 hundreds of feet]

[50,*HMH*: 0.3, 2.0, 4.6, 11.2, 20.3 thousands of feet]

Check is 22+5+23 = 50. Note inclusion of "to" only twice.

The summary of panel E

[50,*HMH*: 3,20,46,112,203]

could also be interpreted as just another trail off toward higher values. As we know, this batch is clearly separated into 3 groupings. As we have said, the sort of summary we are making should not detect this separation. Clearly, in this example it does not.

Sometimes we want to show the five numbers in direct comparison with the stem-and-leaf. The easy and effective way to do this is shown in exhibit 5 , where "*3", "H9", "M46", "H32", and "*678" appear in appropriate places between the check count column and the stem column. Notice that dealing with mixed leaves is not a complication. This exhibit shows a less distinctive pattern, where extremes, hinges, and median come quite close to telling the tale. The condensed notation below the stem-and-leaf summarizes the situation quite well. We could -- probably equally well -- have rounded the values further to

[50, *HMH*: 0.2, 1, 2.5, 4.3, 17 millions]

which would remind us that:

- 50 states are considered.
- the smallest state was about 200,000 population in 1960.

exhibit 5 of chapter 2

The 1960 populations of the 50 states

A) STEM-and-LEAF -- unit = 10,000 people (rounded)

		0	
		1	
2	*3	2	39
4		3	93
5		4	5
		5	
11		6	377138
		7	
13	H9	8	69
15		9	75
to 21		1	30,79,74,41,77,86,
✓ 7	M46	2	54,76,18,18,33,38,85,
to 22		3	27,94,04,26,10,41,57,97,95,
13	H32	4	95,66,32,56,
9		5	15,
8		6	07,
7		7	82,
.		8	
6		9	71,58,
4	*678	1	572,008,678,132,
		2	

[50,*HMH*: 0.23, 0.89, 2.5, 4.3, 17 millions]

B) CONVERSION ROUNDING

*:	23	ten thousands	=	0.23	millions
H:	89	" "	=	0.89	"
M:	246	" "	=	2.46	"
H:	432	" "	=	4.32	"
*:	1678	" "	=	16.78	"

C) SOURCE

The World Almanac, 1966, p. 325.
Their source: U. S. Bureau of the Census.

- about half the states fell between one million and 4.3 million -- with about a quarter below 1 and a quarter above 4.3.
- about half the states were smaller and about half larger than 2.5 million.
- the largest was about 17 million.

These reminders really give us a fairly good feel for this whole batch of 50 populations.

The populations of the states clearly trail off much farther above the middle than below. In this instance, this keeps us from condensing this summary still farther, and is something we will need to be able to remedy.

2E. Some Other Convenient Numbers

The five-number summary is quite effective in giving a general impression of the batch. Sometimes, however, we want to be a little more explicit about how spread out the batch is, or at what level its body seems to lie. For spread we can often make good use of

midspread = difference between hinges

which we will mark "dh" to avoid too many marks beginning with the letter m, or

range = difference between extremes

For level we can use the median, but we rather often find advantages in using the

$$\text{trimean} = \frac{\text{hinge} + 2(\text{median}) + \text{hinge}}{4}$$

which we mark "tri" for short.

For exhibit 5, for example, we have

$$\text{midspread} = 3.3 \text{ millions}$$
$$= 4.3 - 1.0$$

$$\text{range} = 16.5 \text{ millions}$$
$$= 16.7 - 0.2$$

$$\text{trimean} = 2.6 \text{ millions}$$
$$= \frac{1.0 + 2(2.5) + 4.3}{4}$$

For panel E of exhibit 4, similarly,

$$\text{midspread} = 9.2 \text{ thousand feet}$$
$$= 11.2 - 2.0$$

$$\text{range} = 20.0 \text{ thousand feet}$$
$$= 20.3 - 0.3$$

$$\text{trimean} = 5.6 \text{ thousand feet}$$
$$= \frac{2.0 + 2(4.6) + 11.2}{4}$$

In this latter instance, the trimean is appreciably different from the median, reflecting the dissymmetry of the batch of heights.

If, as we said above, our concern is with the level of the body of the batch -- not just with its center -- the trimean often serves us better than the median.

A convenient condensed notation that we will use freely is illustrated -- for the examples of panels D and E just written of exhibit 4 -- by:

82			50	
H	45		H	2.0
M	57	and	M	4.6
H	69		H	11.2
tri	57		tri	5.6

To reach these figures we had to work out the depths of both hinges and median. Writing them down in a standard way is quite helpful, and makes checking easier. The recommended form is:

82			50	
21H	45		13H	2.0
41.5M	57	and	25.5M	4.6
21H	69		13H	11.2
tri	57		tri	5.6

We are also well aware that the level of batches is sometimes summarized, for better or worse, by the mean of its values.

2F. Pentameans?

Trimeans are effective summaries. They use 3 of the 5 summary values that we have used as a group. Should we use the other 2 summary numbers -- the extremes -- also?

Whether we stare at one batch after another of real data summarized by 5-number summaries -- or whether we try to develop a rather sophisticated mathematical description of how data might be -- we find that we could take some account of the values of the extremes, but that we

|| ought not take too much account of the extremes. ||

Try a few batches, made up out of whole cloth or modified from real data, and learn to what extent you agree with this result.

If we wanted to bring in the extremes, we could perhaps use such a summary as:

$$\text{pentamean} = \frac{\frac{1}{10}(\text{extreme})+(\text{hinge})+(\text{median})+(\text{hinge})+\frac{1}{10}(\text{extreme})}{3.2} .$$

On the whole, this seems rarely worthwhile. There is rarely much to gain, sometimes there is much to lose. Always there must be appreciably more effort.

Accordingly we shall not try to calculate or use pentameans.

The role of the extremes, in a summary is, in fact, not as further help in telling us where the values are located or centered. Rather they help us by telling us how far the stray values do stray.

Once we recognize this as the role of the extremes, we are happy to go ahead -- as a routine matter -- choosing between the trimean and the median in most, if not all, instances.

2G. Halving and Trimeans (helpful)

The working data analyst cannot avoid halving and doubling. These arithmetic steps should become automatic.

When calculating a trimean mentally, we should do two halvings, thinking of

$$\frac{\text{hinge} + 2(\text{median}) + \text{hinge}}{4}$$

as if it were written

$$\frac{\left(\frac{\text{hinge} + \text{hinge}}{2}\right) + \text{median}}{2}$$

Thus in the last example of panel E, exhibit 4 with

[HMH: 2.0,4.6,11.2]

we would calculate mentally (or sotto voce):

2.0 and 11.2 is 13.2,
half 13.2 is 6.6,
4.6 and 6.6 is 11.2,
half 11.2 is 5.6.

The experienced halver, however, would jump from "4.6 and 6.6" to "5.6" as obviously halfway between the given values.

The order of calculation can affect rounding, thus

[HMH: 2,5,17]

leads to a trimean of 8 by this route

$$\frac{2+17}{2} = 10$$

$$\frac{5+10}{2} = 8$$

while doing the calculation in the stiff formal way would give

$$\frac{2 + 2(5) + 17}{4} = 7$$

or

$$\frac{2 + 2(5) + 17}{4} = 7.2$$

Don't be surprised, especially as you read on, to find an "8" as an answer in this situation.

If it mattered which we got, we probably ought to keep the extra digit and get 7.2, either as above or, better, this way:

$$\frac{2+17}{2} = 9.5$$

$$\frac{5+9.5}{2} = 7.2$$

Quite often, it doesn't matter.

*****

2H. <u>How far have we come?</u>

This chapter has been concerned with the simplest numerical summaries for a batch of numbers. We have considered two stages of summarization:

- picking out 5 numbers from among those observed that will tell us about the general behavior of the batch.
- focussing down on either or both of
  - -- a central value to tell us of the level of the body of the batch
  - -- a measure of its spread.

Picking out values from the batch is made simpler by the use of ranks, defined by counting in from either end, the extreme observations being given rank 1. Interpolated ranks are also a help, with the value given rank 4.5, for instance, being the mean of the values given ranks 4 and 5.

Depth, the lesser of the two ranks for a given value is a natural and central notion.

The notion of "grows to", mentioned in the introduction to chapter 1, is also useful. Any number not an integer "grows to" the next larger integer. Any integer grows by 1/2.

The five numbers we selected as the first stage of summarization were:

- the two extremes, with depth = 1. (Marked "*".)

- the two hinges, the numbers whose depth is what "count/4" grows to. (Marked "H".)

- the median, the number whose depth is what "count/2" grows to. (Marked "M".)

Written down in convenient condensed form -- a choice between two was offered -- these five numbers offer as a good description of the more routine aspects of the pattern of the batch. As we have wished, they do not try to tell us about more unusual aspects of the batch's pattern. We still have to depend on a stem-and-leaf display for such information.

At the second stage of summarization we mentioned two central values that we might consider:

- the median -- specified above.

- the trimean -- one-quarter the sum of the two hinges and twice the median.

(A later optional section will mention two more: the midmean and the mean.)

If we want our central value to tell us about the general level of the body of the batch, which is a common case, neither the median -- which looks at too few values -- or the mean -- which looks at too many -- is satisfactory. The trimean is often a wise choice in such circumstances.

Two measures of spread were mentioned:

- the midspread, which is the difference between the hinges. (Marked "dh".)
- the range, which is the difference between the extremes.

Arithmetic details in the calculation of trimeans were also considered, as was the possible use of pentameans.

****

2K. Middle Means

It takes very little experience to learn that -- when stem-and-leaf is used -- medians are much less work than means, and that trimeans are but little more work than medians. Moreover, hinges are worthwhile for their own sake. So trimeans need to be taken very seriously.

What we lack, however, is a framework in which to place the trimean, a framework that will let us think about it easily and effectively.

Both median and trimean differ from the mean in paying no attention to the values of the numbers in the end quarters -- in the wings. This is often a good thing, as we shall see. When we ask about "level" it is often the "body" of our numbers that concerns us, not all of them.

In any batch, those most subject to laying aside are the end numbers. The median comes about, in one interpretation, by laying aside some of the highest numbers and some of the lowest numbers, almost 1/2 of all numbers in each case. A reasonable compromise between doing this and laying aside nothing, as the mean does, is to lay aside about 1/4 of all the numbers at each end. The easiest thing to do with what then is left is to take a mean of either all or some of the middle numbers that then remain. Since any such result is a mean of observations taken from the middle of the batch, it is natural to call it a

middle mean

and to require that it "behave" rather like the mean of all or most of these middle numbers. (This last condition rules out the median of a large batch, which is a mean of one or two middle numbers but fails to behave like the mean of most of the middle numbers.)

We wish to be flexible, rather than rigid, so we include the hinges with the numbers ranked between the as proper candidates for forming a middle mean. This makes the trimean a middle mean -- if it behaves like one. That it does so is not obvious. Experience, however, of which any reader can gain some by simple trials, reassures us that it does behave properly -- that it is a middle mean.

Its natural competitor is the

midmean

the mean of all the numbers ranked between the hinges. Even in this brief mention, we need to notice that,

for batches of count 2 to 6

midmean = median

while the trimean only reduces to the median in batches of 1 and 2.

We shall come back to the choice between trimean and midmean in a later chapter. For the present we can be ready to use either one.

2L. <u>Why might the trimean behave like a middle mean?</u>
(optional background)

Given a large batch, median and hinges divide it into four quarters. The mean of all the observations in the two middle quarters is -- by definition -- a middle mean. If we want a quick approximation, it helps to think of this as the mean of two numbers -- of the means of the individual quarters.

To find the quarters we must find their boundaries -- the hinges and the median. An easy, rough replacement for the mean of all the observations in a quarter is the mean of its two boundaries. There will be a tendency for this result to be somewhat farther away from the median than we should like. Since these shifts are typically outward for both quarters -- and hence tending toward mutual cancellation so far as the mean of these results is concerned, we might expect this mean which we can write

$$\frac{1}{2}\left(\frac{\text{hinge + median}}{2}\right) + \frac{1}{2}\left(\frac{\text{median + hinge}}{2}\right)$$

$$= \frac{\text{hinge + 2(median) + hinge}}{4} = \text{trimean}$$

to be a middle mean.

****

2P. Problems for chapter 2.

1. Summarize in 5 numbers the areas of the counties in the 5 states given in exhibit 6. Comment on the similarities and differences.

2. The German atomic weights for 1921 were based on the determinations listed in exhibit 7. Summarize the results for each element. What are the midspreads and ranges? Do they seem to be related to the medians? Any comment?

3. Identify all observations of depth 2 in exhibit 7.

4. According to the 5th edition of Landolt-Bornstein's (1923) the available careful determinations of the weight in grams of a liter of oxygen gave these values: 1.42905, 1.42906, 1.42894, 1.42876, 1.42896, 1.42915, 1.4290, 1.42905, 1.42906. Make an appropriate stem-and-leaf display. Find the five-number summary. State midspread and range.

5. In 1950, Goldberg and Brown gave the batches of measured gold contents of meteorites given in exhibit 8. Make five-number summaries. Find midspreads and ranges. Comments on results.

exhibit 6 of chapter 2

Sizes of counties in five states

A) The AREAS -- unit = 100 square miles

Arizona

1	92
2	
3	
4	86
5	4
6	3
7	
8	1
9	292
10	0
11	2
12	
13	3

More:186

Connecticut

0	67936745

Rhode Island

0	02143

Maine

0	5994537
1	750
2	16
3	499

More:68

Massachusetts

0	4961576580461
1	5

B) The SOURCE

<u>The World Almanac</u> and 1966. Page 363 and thereafter.

exhibit 7 of chapter 2

Atomic weight determinations

used in fixing the German atomic weights for 1921

A) The DATA

Argon (page 2)	**Copper** (page 4),	**Iodine** (page 6)
39.88	63.57	126.924
39.91	63.576	126.925
39.88	63.572	126.923
39.89	63.53	126.938
39.9	63.56	126.932
39.945	63.57	126.913
39.91	63.57	126.925

Antimony* (page 9)	**Antimony**** (page 9)	**Tellurium** (page 11)
120.55	122.96	127.53
120.55	121.83	127.62
119.86	121.24	127.53
119.79	122.08	127.57
120.35	122.37	127.54
	121.77	127.479
		127.6
		127.51
		127.8

B) The SOURCE

(Landolt-Bornstein, vol. 1)

exhibit 8 of chapter 2

Parts per million of gold in meteorites

A) The DETERMINATIONS -- 4 meteorites

Canyon Diablo		Henburs		Xiquipitco		Odessa	
2.6	6	0.6	65,92	2.2	5	2.2	4
2.8	48	0.7	00,01,57,74	2.3	14	2.3	1
2.9	0	0.8	09	2.4	4	2.6	3
3.0	023	0.9	10,36,54	2.5	7	2.7	24
3.1	39			2.6	034	2.8	2
3.2	2			2.7	46	2.9	1
				2.8	0	---	
				---		3.1	
				3.1	0	3.2	

						3.6	

B) The SOURCE

E. D. Goldberg and H. Brom 1950. Radiometric determination of gold and rhenium. Analytical Chemistry 22: 308-311. Table 1 on page 310.